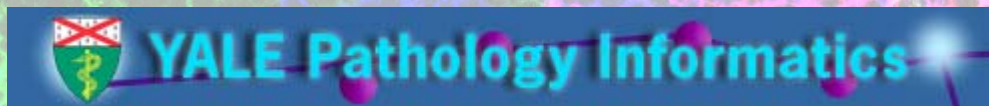


# Modeling & Simulation (Computational Immunology)

**Steven H. Kleinstein**



Departments of Pathology and Immunobiology  
Yale School of Medicine

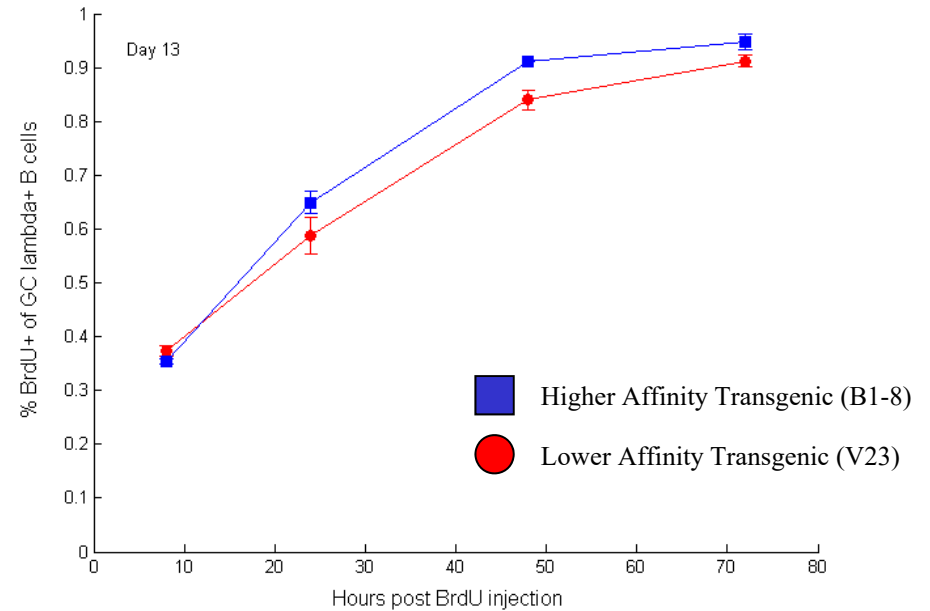
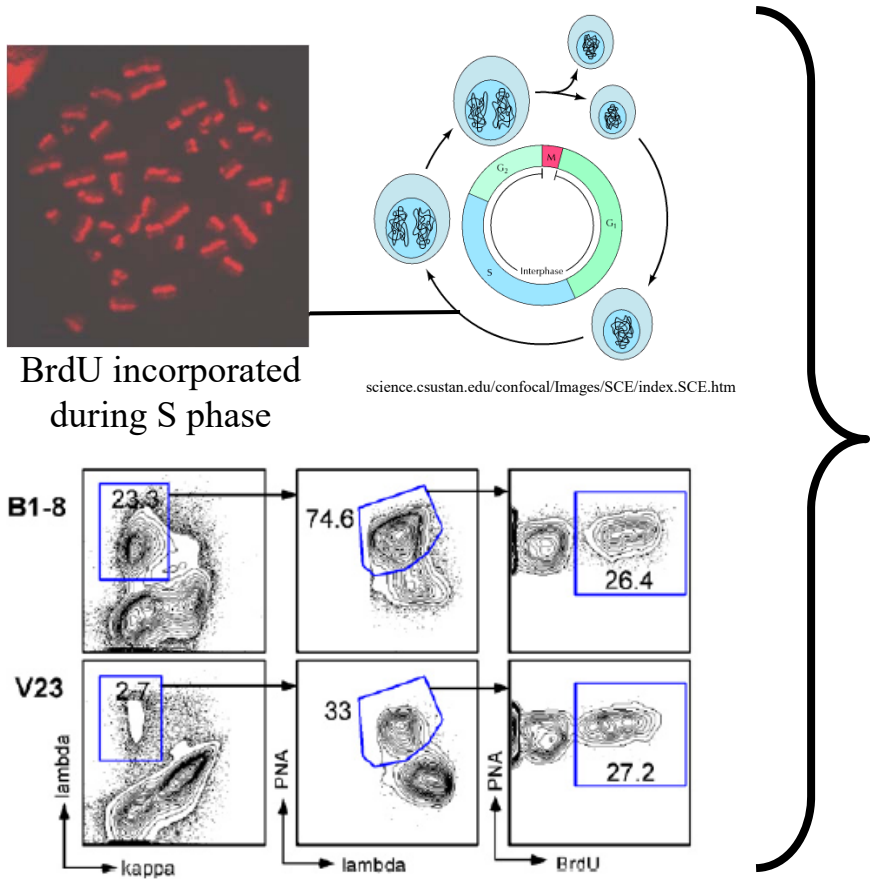
[steven.kleinstein@yale.edu](mailto:steven.kleinstein@yale.edu)

March 25, 2020

# Understanding cell proliferation and death

BrdU (thymidine analog) incorporated into cell DNA during S-phase

Flow cytometry to quantify antigen-specific germinal center B cells...



Labeling curves look similar – suggests same proliferation rate?

# Understanding cell proliferation and death

At steady-state, rate at which the fraction of BrdU labeled cells increases is indicative of the sum of the per cell proliferation and death rates

## Quantification of Cell Turnover Kinetics Using 5-Bromo-2'-deoxyuridine<sup>1</sup>

Sebastian Bonhoeffer,<sup>\*</sup> Hiroshi Mohri,<sup>†</sup> David Ho,<sup>†</sup> and Alan S. Perelson<sup>2,\*‡</sup>

*The Journal of Immunology*, 2000, 164: 5049–5054.

## Rapid Turnover of T Lymphocytes in SIV-Infected Rhesus Macaques

Hiroshi Mohri, Sebastian Bonhoeffer, Simon Monard, Alan S. Perelson, David D. Ho<sup>\*</sup>

www.sciencemag.org • SCIENCE • VOL. 279 • 20 FEBRUARY 1998

*International Immunology*, Vol. 15, No. 3, pp. 301–312  
doi:10.1093/intimm/dxg025, available online at www.intimm.oupjournals.org

## Asynchronous differentiation models explain bone marrow labeling kinetics and predict reflux between the pre- and immature B cell pools

Ramit Mehr<sup>1</sup>, Gitit Shahaf<sup>1</sup>, Alex Sah<sup>2</sup> and Michael Cancro<sup>2</sup>

The Journal of Immunology

## Taking Advantage: High-Affinity B Cells in the Germinal Center Have Lower Death Rates, but Similar Rates of Division, Compared to Low-Affinity Cells<sup>1</sup>

Shannon M. Anderson,<sup>\*</sup> Ashraf Khalil,<sup>†</sup> Mohamed Uduman,<sup>§§</sup> Uri Hershberg,<sup>\*†§</sup> Yoram Louzoun,<sup>¶</sup> Ann M. Haberman,<sup>‡</sup> Steven H. Kleinstein,<sup>§§</sup> and Mark J. Shlomchik<sup>2\*†</sup>

*Oncogene* (2005) 24, 7514–7523

© 2005 Nature Publishing Group. All rights reserved 0950-9232/05 \$30.00

www.nature.com/onc

## Reduced cell turnover in lymphocytic monkeys infected by human T-lymphotropic virus type 1

Christophe Debacq<sup>1,5</sup>, Jean-Michel Héraud<sup>2,5</sup>, Becca Asquith<sup>3</sup>, Charles Bangham<sup>3</sup>, Fabrice Merien<sup>2</sup>, Vincent Moules<sup>4</sup>, Franck Mortreux<sup>4</sup>, Eric Wattel<sup>4</sup>, Arsène Burny<sup>1</sup>, Richard Kettmann<sup>1</sup>, Mirdad Kazanji<sup>2</sup> and Luc Willems<sup>\*.1</sup>

Models of BrdU incorporation integral part of many studies

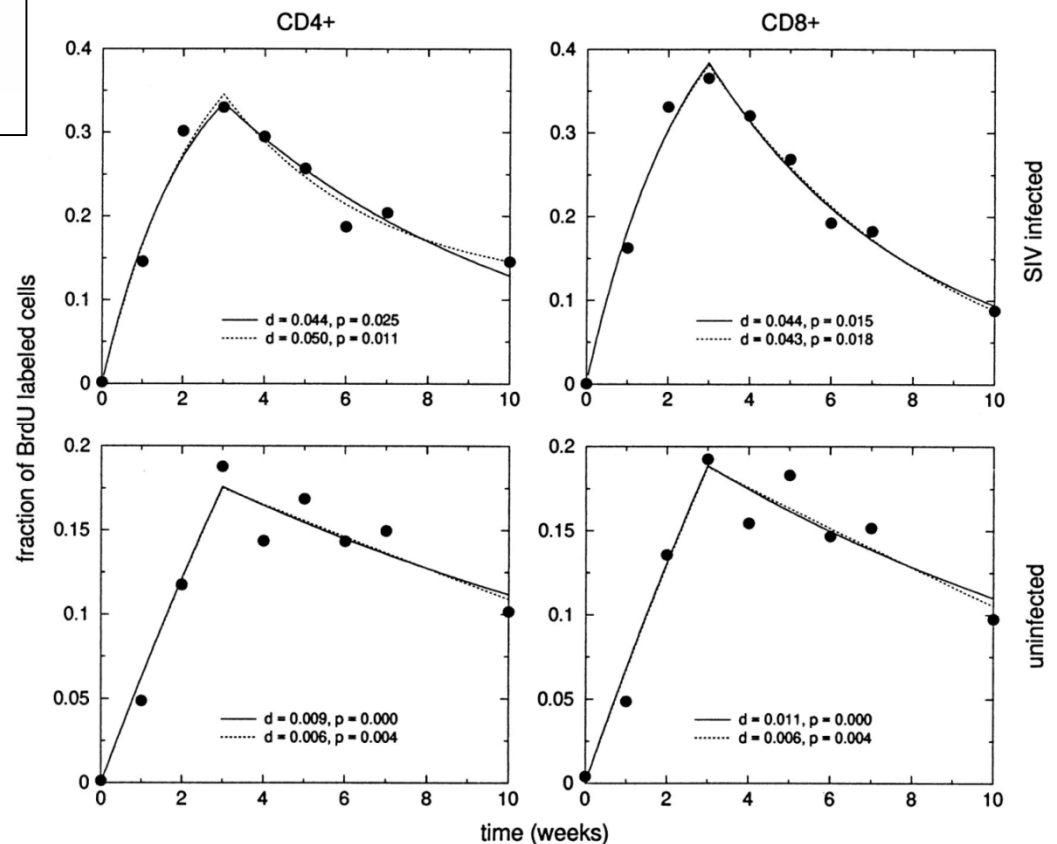
# BrdU labeling of CD4+ and CD8+ T lymphocytes

SIV-infected and an uninfected macaque. Data are from Mohri et al., Science (1998)

## Rapid Turnover of T Lymphocytes in SIV-Infected Rhesus Macaques

Hiroshi Mohri, Sebastian Bonhoeffer, Simon Monard, Alan S. Perelson, David D. Ho\*

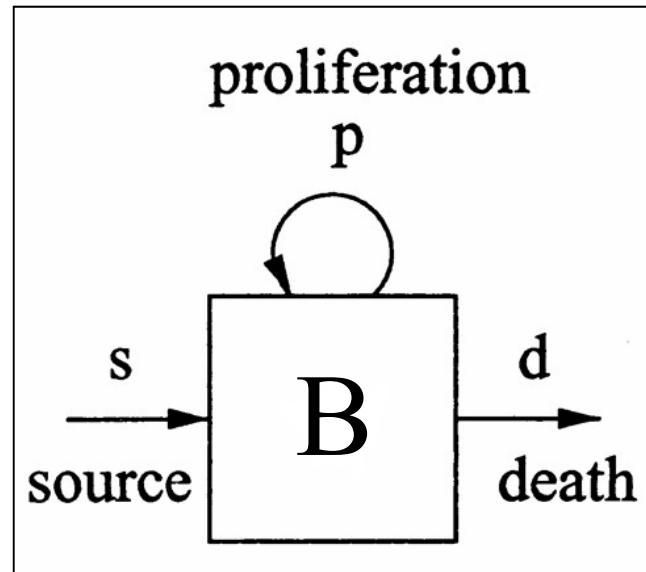
www.sciencemag.org • SCIENCE • VOL. 279 • 20 FEBRUARY 1998



Is there a difference in cell turnover?

# Model of BrdU Labeling

Start with a basic model of cell population dynamics...



Rate of change  
in B cell  
population

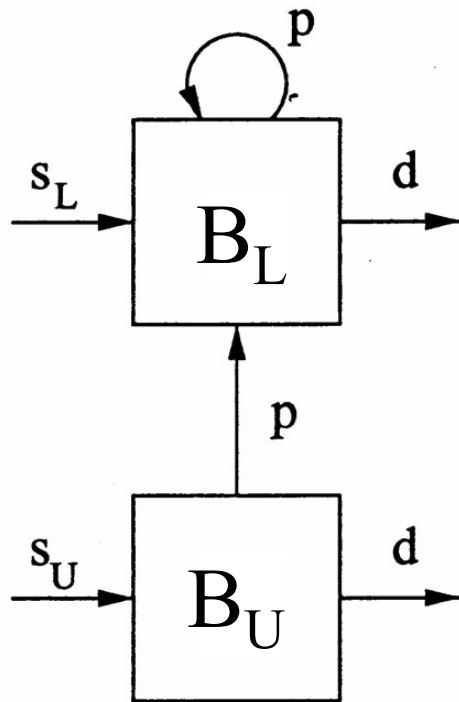
$$\frac{dB}{dt} = s + pB - dB$$

Often can often assume population in steady-state (i.e., constant)

# Model of BrdU Labeling

Split the B cell population into Labeled ( $B_L$ ) and Unlabeled ( $B_U$ ) subsets

B) During BrdU administration



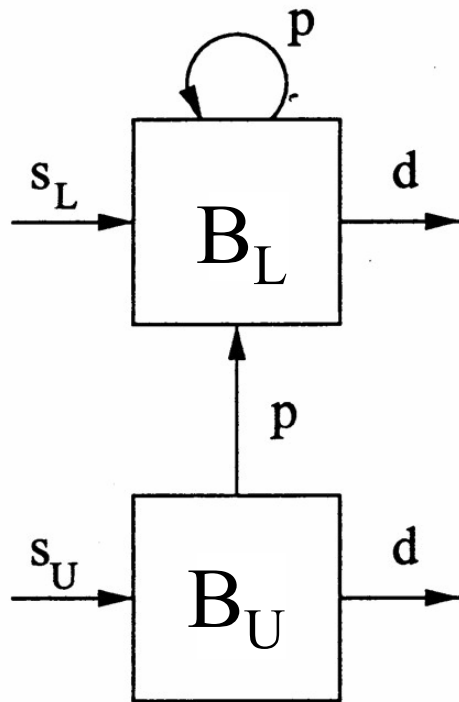
$$\frac{dB_U}{dt} = s_u - pB_U - dB_U$$
$$\frac{dB_L}{dt} = s_l + 2pB_U + pB_L - dB_L$$

Do data contain enough information to estimate parameters?

# Model of BrdU Labeling

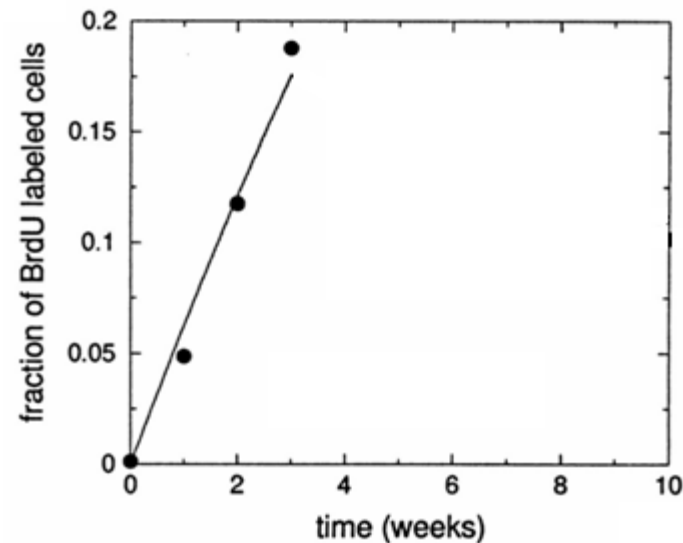
Label is administered continuously over some time-period

B) During BrdU administration



$$f_L(t) = A_1 (1 - e^{-(d+p)t})$$

$$A_1 = 1 - \frac{s_U}{(s_U + s_L)} \times \frac{(d - p)}{(d + p)}$$



Labeling curve reflects both proliferation AND death

# Model Identifiability

A model is identifiable if possible to learn true value of underlying parameter after obtaining enough observations

**Identifiable parameters** are those which effect the value of the data and can be estimated with some degree of certainty.

**Non-identifiable parameters** are those which effect the value of the data but which cannot be estimated accurately

**Non-observable parameters** are those which don't have an effect on the data.

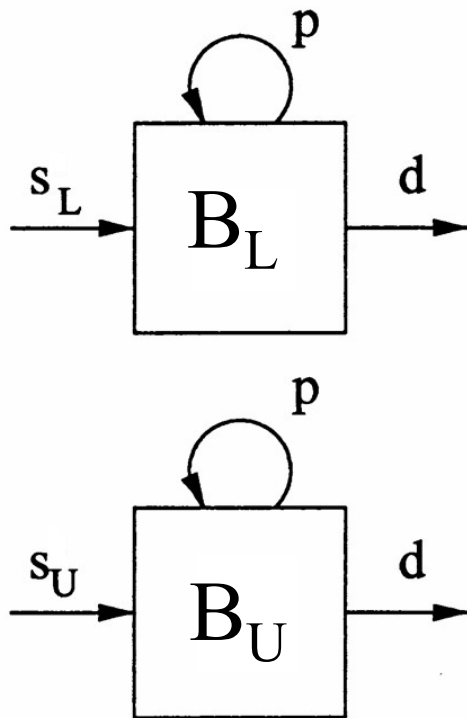
Cannot estimate both proliferation AND death



# Model of BrdU DE-Labeling

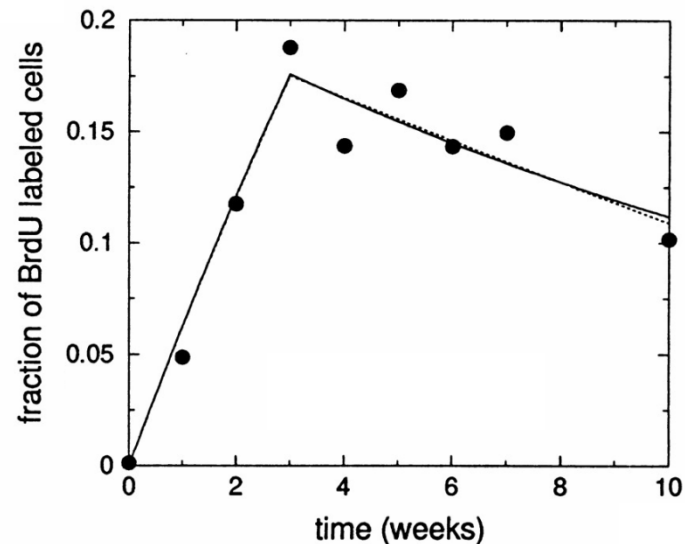
Stop administering label after some time ( $t_e$ )

C) After BrdU administration



$$f_L(t) = A_2 + A_3 e^{-(d-p)(t-t_e)}$$

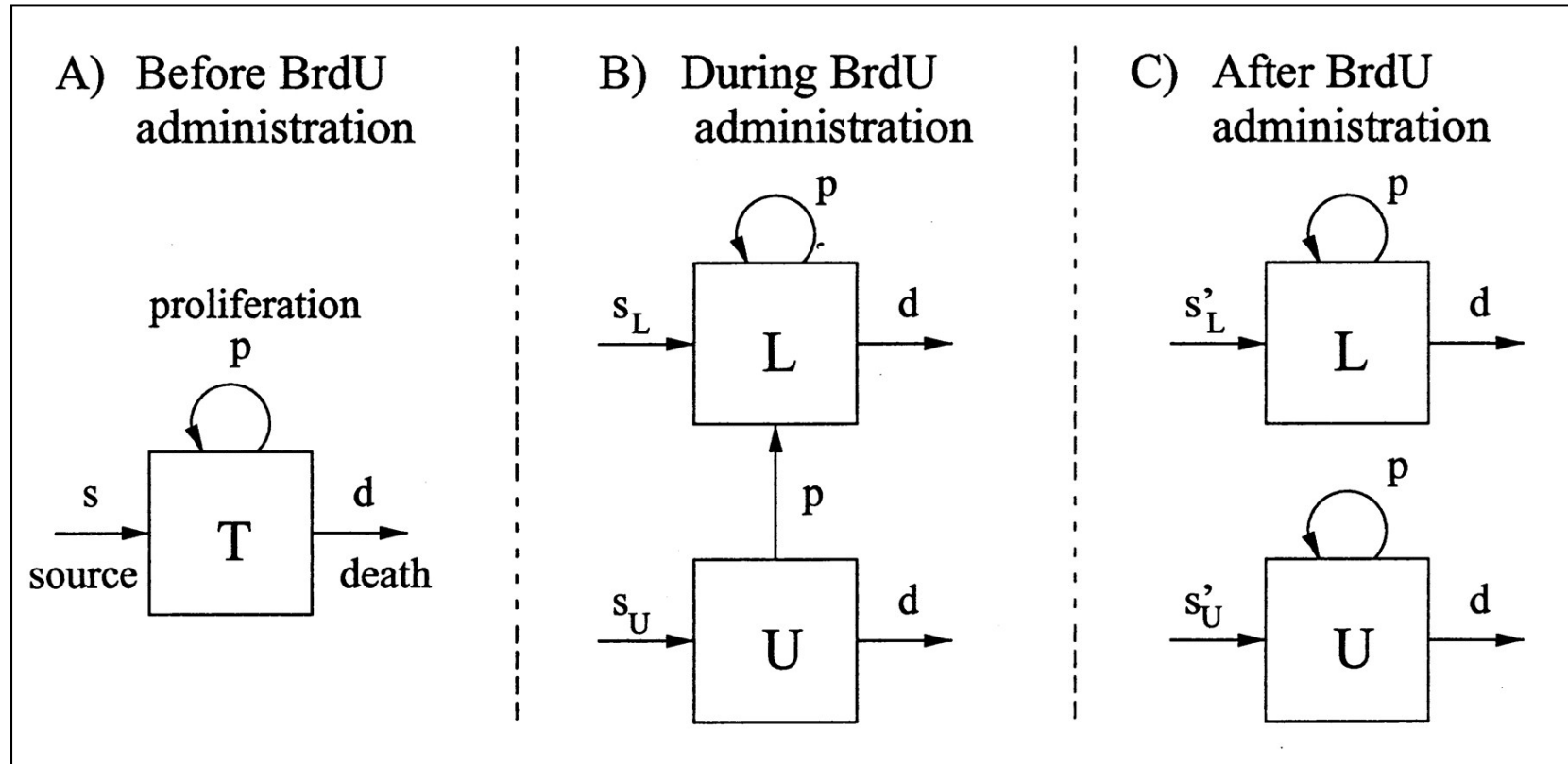
$$A_2 = \frac{s_L}{(s_U + s_L)}, A_3 = f_L(t_e) - A_2$$



Now, we can estimate BOTH proliferation AND death

# Model of BrdU Labeling

Model changes with experiment



We can express these as sets of ordinary differential equations

# Characteristics of a Good Inverse Model

- Fit is good—model should be able to adequately describe a relatively noise-free data set (of course a poor fit provides some insight also)
- Model parameters are unique
  - Theoretically identifiable for noise-free data
  - Well-determined model parameters in presence of measurement noise
- Values of parameter estimates are consistent with hypothesized physical or physiologic meanings and change appropriately in response to alterations in the actual system

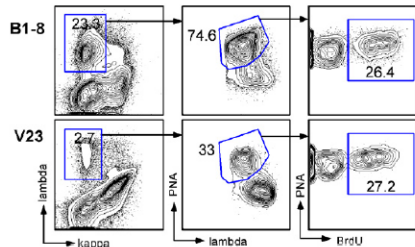
# Six Steps for Inverse-Modeling of Data

1. Select an appropriate mathematical model
  - Polynomial or other functional form
  - Based on underlying theoretical equations
2. Define a “figure of merit” function
  - Measures agreement between data & model for given parameters
3. Adjust model parameters to get a “best fit”
  - Typically involves minimizing the figure of merit function
4. Examine “goodness of fit” to data
  - Never perfect due to measurement noise
5. Determine whether a much better fit is possible
  - Tricky due to possible local minima vs. global minimum
  - F-test for comparing models of different complexity
6. Evaluate accuracy of best-fit parameter values
  - Provide confidence limits and determine uniqueness
  - Assess physical reasonability of estimated parameter values

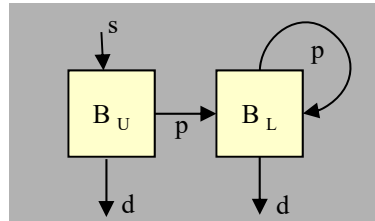
# Interaction of Computation & Experiment

Compare simulation and experiment using least-squares objective

## Experimental Observations



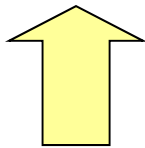
## Computational Model



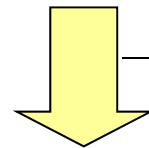
Least-squares objective function

$$E = \sum_i \frac{(y_i - \hat{y}_i)^2}{VAR(y_i)}$$

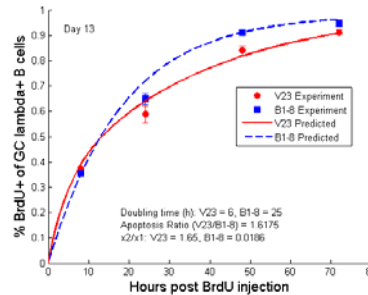
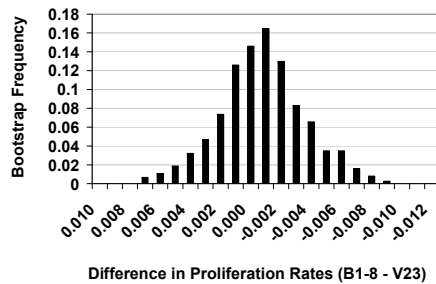
## New Experiments



## Fit Model to Data



## Model Predictions

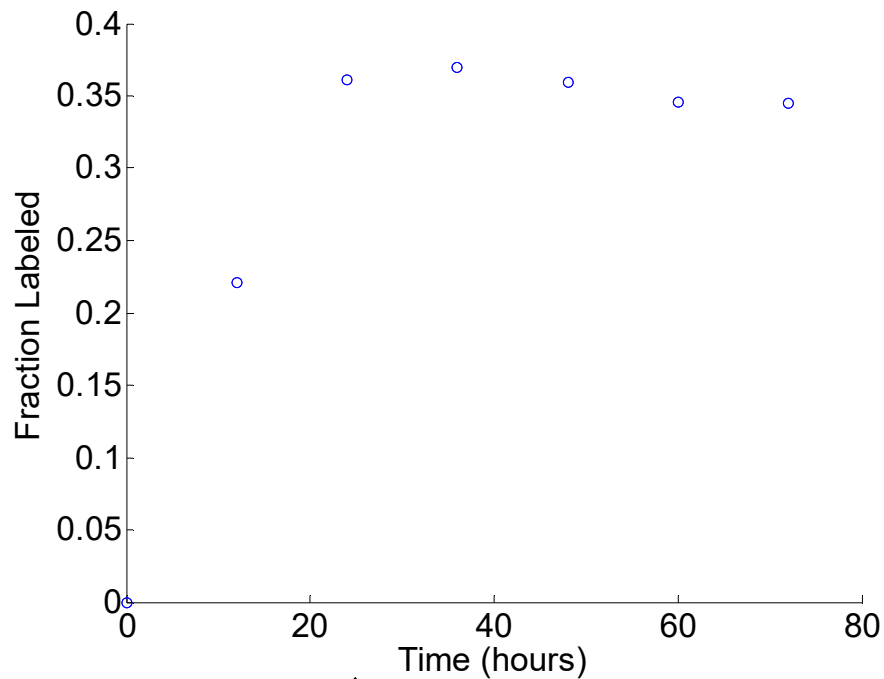


Bootstrapping Confidence Intervals

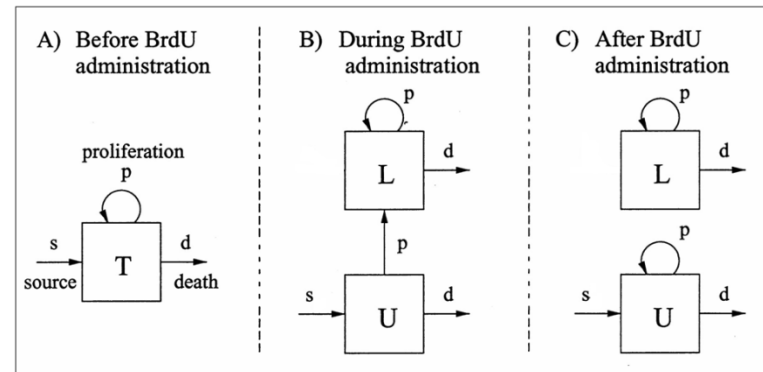
Continuous cycle of modeling and experimentation

# Simulated Experiment

Demonstrate full cycle of fitting model to data to estimate parameters



BrdU withdrawn



## Parameters used to create synthetic data

$s = 0.003$  per hour

$p = 0.01$  per hour

$d = p + s$  (to achieve steady state)

Random noise added to each data point

How can we estimate flow/proliferation/death rates?

# Numerical solution to ODEs

## Euler's Method

$$y'(t) = f(t, y(t)), \quad y(t_0) = y_0,$$

$$y'(t) \approx \frac{y(t+h) - y(t)}{h},$$

$$y(t+h) \approx y(t) + hf(t, y(t)).$$

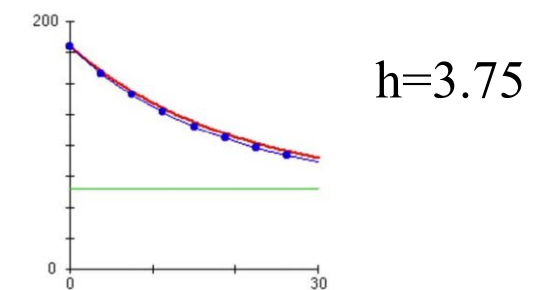
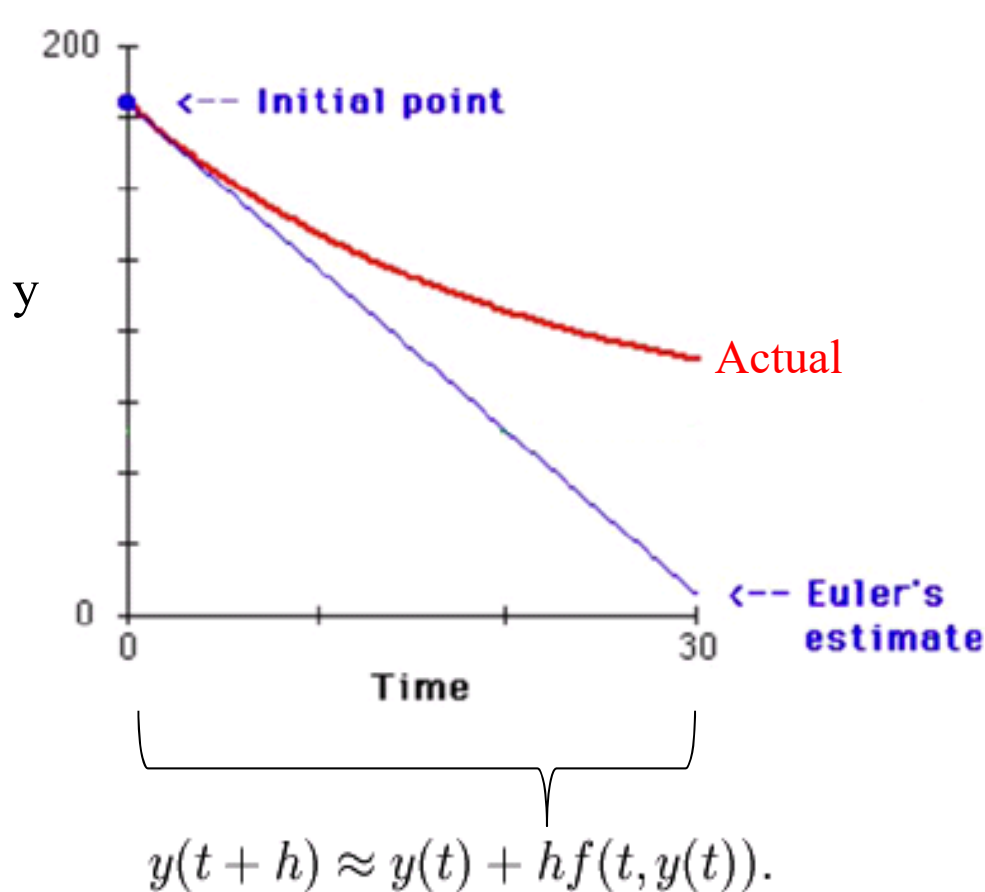


Leonhard Euler  
(1707-1783)

From any point on curve, find approximation of nearby point on curve by moving a short distance along a line tangent to the curve

# Numerical solution to ODEs: Euler Method

From any point on curve, find approximation of nearby point on curve by moving a short distance along a line tangent to the curve

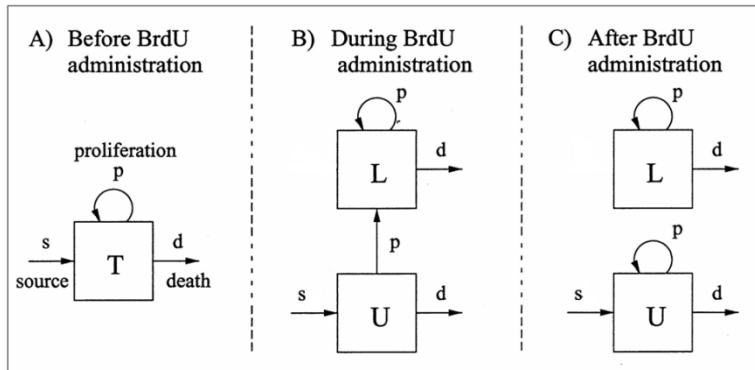


Much better ways to do this in practice. Eg, Runge-Kutta



# Simulating the BrdU Labeling Model

Use integration functions (e.g., ode45 in MATLAB)



```
Yin = [1 0]; % Initial Conditions [unlabeled labeled]
```

```
pr = [s p d tau]; % Model Parameters
```

```
t = [0,12,24,36,48,60,72]; % Times to evaluate
```

```
[T,Y] = ode45(@fode,t,Yin,opts,pr);
```

```
fl = Y(:,2) ./ sum(Y,2); % Fraction labeled
```

```
function dy = fode(t, y, pr)
```

```
s = pr(1); p = pr(2); d = pr(3); tau = pr(4);
```

```
U = y(1); L = y(2);
```

```
dy = zeros(2,1); % Vector of derivatives
```

```
if (t<tau) % During BrdU Administration (B)
```

```
dy(1) = s - p.*U - d.*U; % dbU/dt
```

```
dy(2) = 2.*p.*U + p.*L - d.*L; % dbL/dt
```

```
else % After BrdU Administration (C)
```

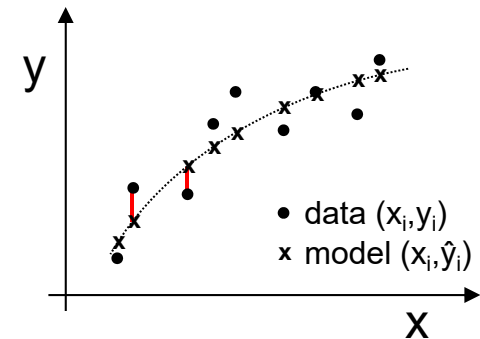
```
dy(1) = s + p.*U - d.*U; %dbU/dt
```

```
dy(2) = p.*L - d.*L; %dbL/dt
```

```
end
```

Simple models can be solved analytically -- faster

# Least-Squares Error Minimization



- Goal is to fit  $N$  data points  $(x_i, y_i) \ i=1..N$
- The model is a function with  $M$  adjustable parameters  $a_k, k=1..M$  used to generate  $N$  model points  $(x_i, \hat{y}_i)$
- The residual measures the difference between a data point and the corresponding model estimate
- Since residuals can be positive or negative, a sum of residuals is not a good measure of overall error in the fit
- A better measure is the sum of squared residuals,  $E$ , which is only zero if each and every residual is zero

$$\hat{y}_i = \hat{y}(x_i, a_1..a_M)$$

$$y_i - \hat{y}(x_i, a_1..a_M)$$

$$\sum_{i=1}^N [y_i - \hat{y}(x_i, a_1..a_M)]$$

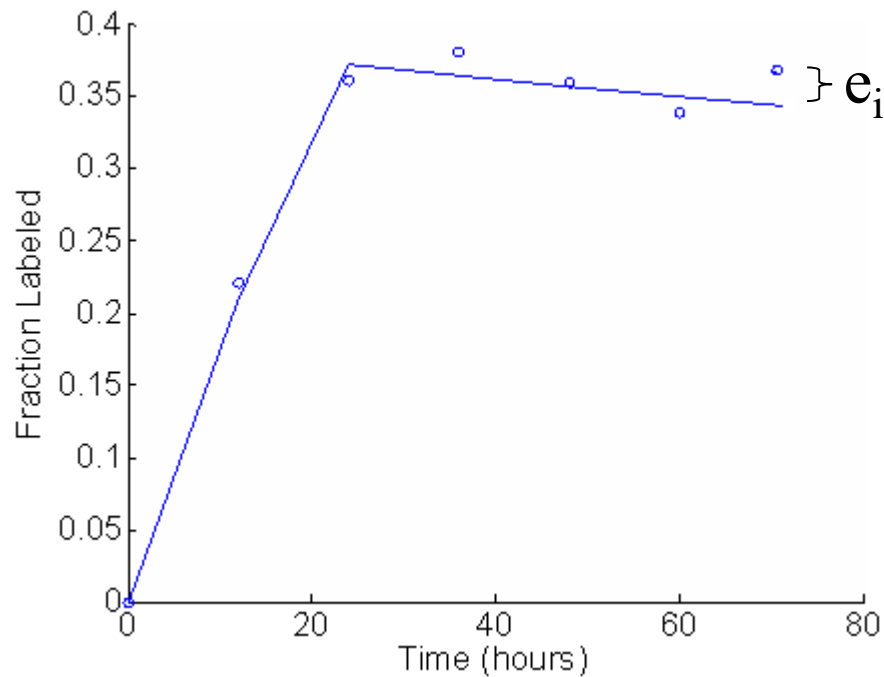
$$E = \sum_{i=1}^N [y_i - \hat{y}(x_i, a_1..a_M)]^2$$

# Maximum Likelihood Estimation

- Not meaningful to ask “What is the probability that my set of model parameters is correct?”
  - Only one correct parameter set → Mother Nature!
- Better to ask “Given my set of model parameters, what is the probability that this data set could be obtained?”
  - What is the likelihood of the parameters given the data?
- Inverse modeling is also known as “maximum likelihood estimation”.

# Fitting the Model to Experimental Data

Compare simulation and experiment using least-squares objective



Least-squares objective function

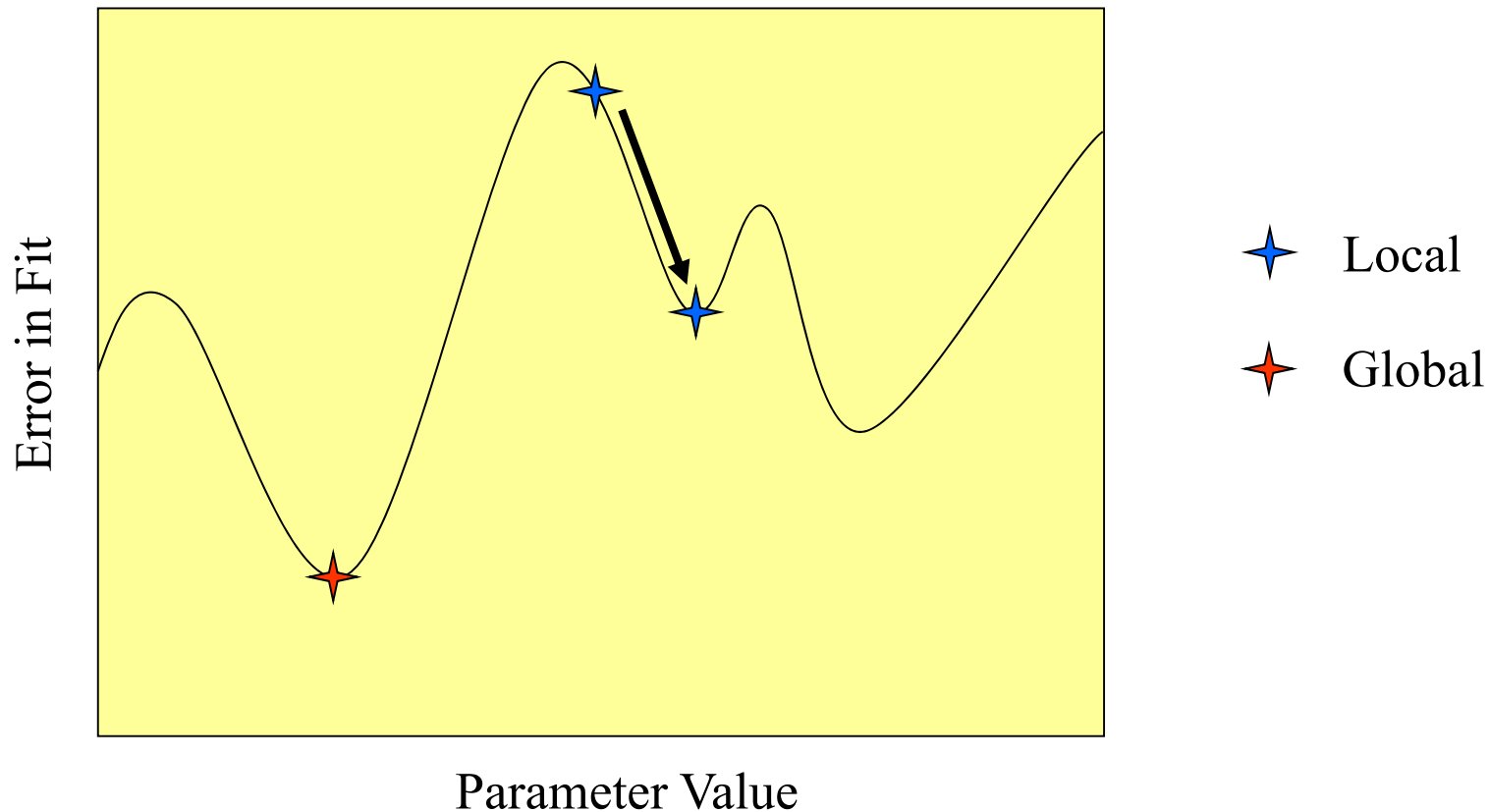
$$E = \sum_i \frac{(y_i - \hat{y}_i)^2}{VAR(y_i)}$$

Find parameters to minimize objective

Many options for how to optimize the fit

# Local and Global Optimization

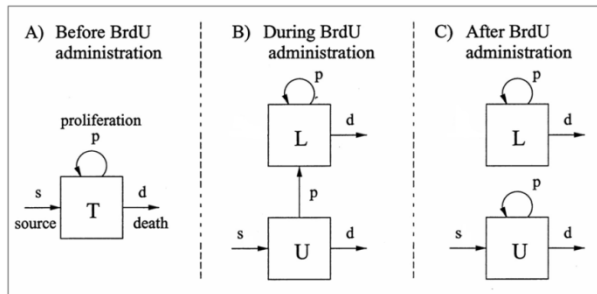
The error function depends on  $M$  model parameters, and can be thought of as an  $M$ -dimensional “surface” of which we seek the minimum



Local optimization techniques find optimal fit around given starting point  
Global optimization attempts to avoid local minima

# Fitting Models to Data in MATLAB

Several optimization functions available in many programming languages



```
pri = [.01 .01]; %Initial guess for parameter values to be fitted [s p]
```

```
[pr,fval,exitflag] = lsqnonlin (@efun,pri,[],[],options,fl_observed,t,tau);
```

```
s = pr(1); p = pr(2); % Optimal parameter values
```

Optional parameters

```
function error = efun (pr,fl_observed,t,tau)
```

```
s = pr(1); p = pr(2); d = s+p; % Assume steady-state
```

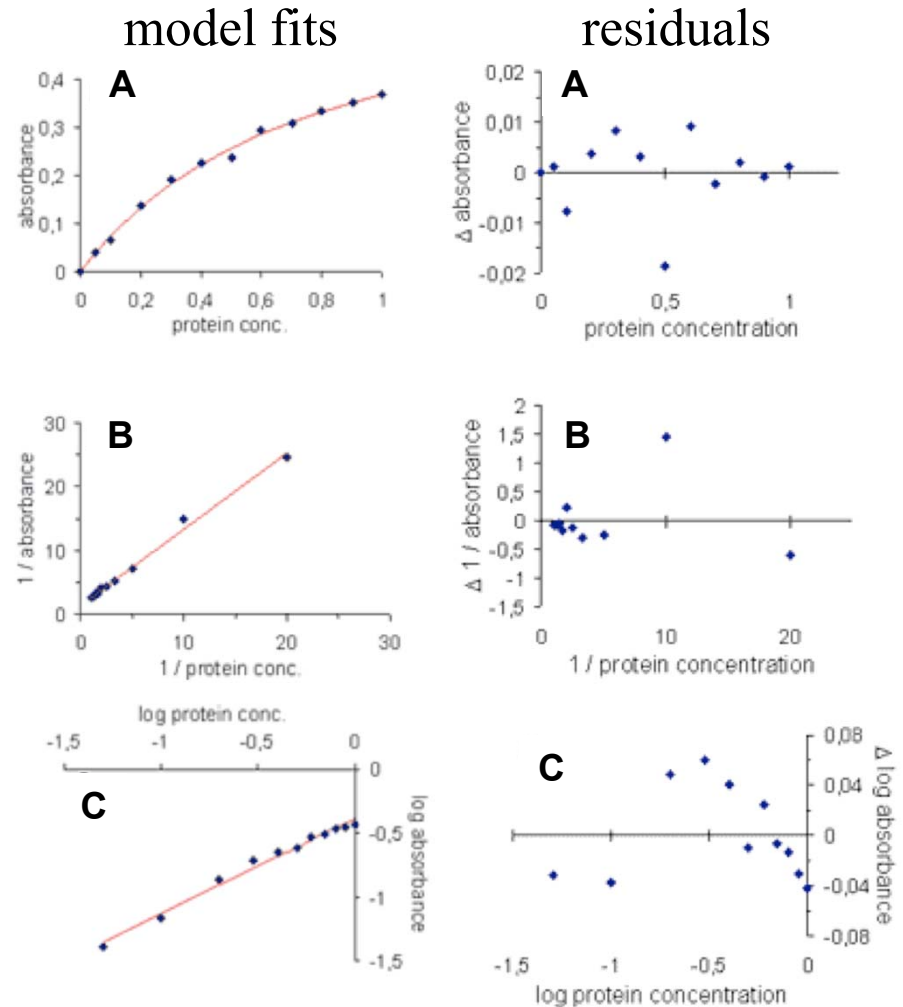
```
[fl_predicted] = labelBrdU(s,p,d,tau,t); % Function that simulates model
```

```
error = sum((fl_predicted-fl_observed).^2); % Least-squares objective
```

lsqnonlin, fminsearch, fmincon, fminbnd

# Goodness of Fit and the Residuals Plot

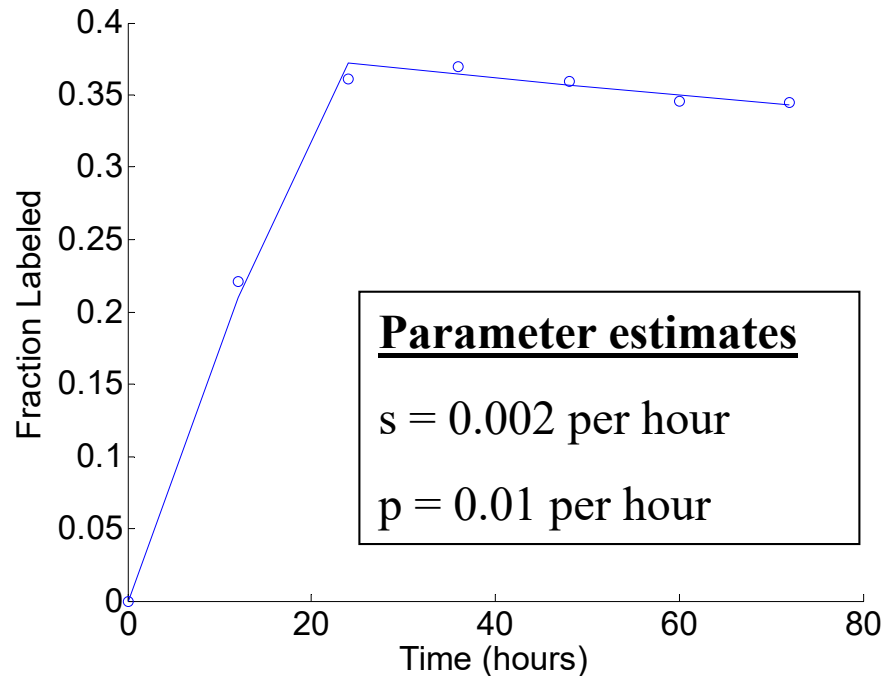
- A high correlation can exist even for a model that systematically differs from the data (all 3 examples have  $r^2 > 0.99$ )
- One must also examine the distribution of residuals—a good model fit should yield residuals equally distributed along x and normally distributed around zero with no systematic trends, as in A rather than B or C



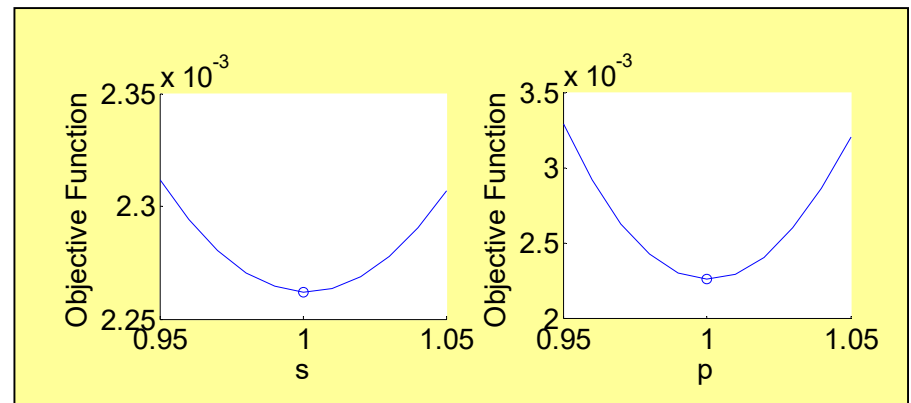
adapted from Lobemeier, 2000

# Optimal Parameter Estimates

Least-squares fit using lsqnonlin in MATLAB



Plot local curvature to check minimization...



**Recall, parameters used to create data:**

$s = 0.003$  per hour

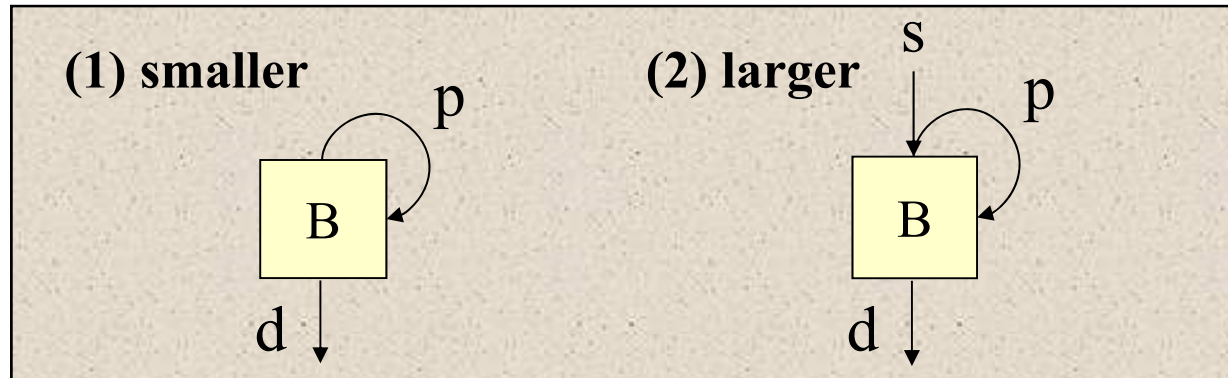
$p = 0.01$  per hour

$d = p + s$  (to achieve steady state)

Is inflow necessary to fit the data? Can we use simpler model?



# Is inflow (s) significant?



Residual Sum of Squares  $RSS = \sum_i (y_i - \hat{y}_i)^2$

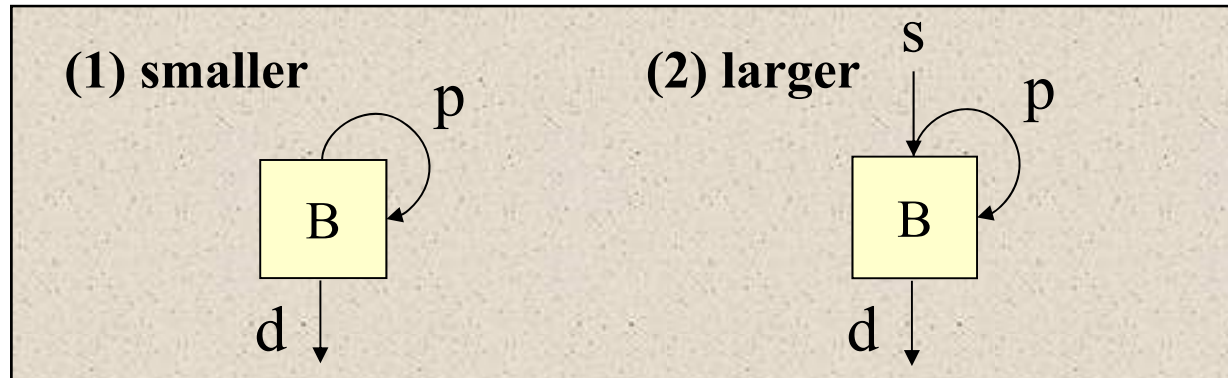
$$F = \frac{RSS_{\text{smaller}} - \cancel{RSS_{\text{larger}}}}{\cancel{RSS_{\text{larger}}} / df_{\text{larger}}} \frac{df_{\text{smaller}} - df_{\text{larger}}}{df_{\text{larger}}}$$

} Reduction in RSS per extra parameter  
 } Measure of 'noise' in model

Degrees of Freedom  $df = \# \text{ observations} - \# \text{ parameters}$

F distribution with  $(df_{\text{smaller}} - df_{\text{larger}}, df_{\text{larger}})$  degrees of freedom

# Is inflow (s) significant?



$$F = \frac{\text{RSS}_{\text{smaller}} - \text{RSS}_{\text{larger}}}{\text{RSS}_{\text{larger}}} \cdot \frac{df_{\text{smaller}} - df_{\text{larger}}}{df_{\text{larger}}}$$

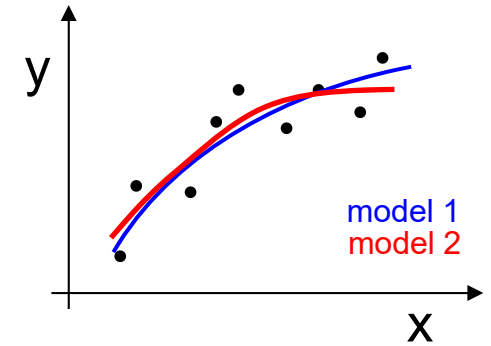
} Reduction in RSS per extra parameter  
 } Measure of 'noise' in model

	Observations	Parameters	RSS	F test (1-fcdf in MATLAB)
<b>(1) No flow (s=0)</b>	6	1	9.38e-7	
<b>(2) Including flow</b>	6	2	0.95e-7	<b>53.1 (p&lt;0.0004)</b>

Inflow (s) is important to explain observations

# Comparing Two Model Fits

- The number of data points,  $N$ , must exceed the number of model parameters,  $M$ , yielding the degrees of freedom ( $DOF = N - M$ )
- Increasing  $M$  using a more complex model will generally improve the quality of fit and reduce RSS
- Increasing  $MSE$  with decreasing RSS can reveal an over-parameterized model
- An F-statistic can be computed to compare the results of two model fits
  - $F \sim 1$ , the simpler model is adequate
  - $F > 1$ , the more complex model is better, or random error led to a better fit with the complex model
  - P-value defines the probability of such a “false positive” result (lookup in F table)



$$M \leq N - 1$$

$$MSE = \frac{RSS}{N - M} = \frac{RSS}{DOF}$$